SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution

Abstract

Owe to the powerful generative priors, the pretrained text-to-image (T2I) diffusion models have become increasingly popular in solving the real-world image super-resolution problem. However, as a consequence of the heavy quality degradation of input low-resolution (LR) images, the destruction of local structures can lead to ambiguous image semantics. As a result, the content of reproduced high-resolution image may have semantic errors, deteriorating the super-resolution performance. To address this issue, we present a semantics-aware approach to better preserve the semantic fidelity of generative real-world image super-resolution. First, we train a degradation-aware prompt extractor, which can generate accurate soft and hard semantic prompts even under strong degradation. The hard semantic prompts refer to the image tags, aiming to enhance the local perception ability of the T2I model, while the soft semantic prompts compensate for the hard ones to provide additional representation information. These semantic prompts encourage the T2I model to generate detailed and semantically accurate results. Further-more, during the inference process, we integrate the LR images into the initial sampling noise to mitigate the diffusion model's tendency to generate excessive random details. The experiments show that our method can reproduce more realistic image details and hold better the semantics. The source code of our method can be found at https://github.com/cswry/SeeSR.